DOCUMENT RESUME

ED 247 270 TM 840 443

AUTHOR Kulik, James A.
TITLE Uses and Misuses of Meta-Analysis.
PUB DATE Apr 84
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Trends; *Meta Analysis; *Research Methodology; Research Problems; Statistical Data
IDENTIFIERS Replication

ABSTRACT

Meta-analysis does not solve all the problems encountered in qualitative research reviews, but it holds out some hope for increasing the reliability and dependability of a reviewer's conclusions. Several developments in meta-analysis are cause for optimism. First, different meta-analysts are doing work in the same areas, leading to increased confidence in meta-analytic results. Second, meta-analysts are beginning to include their raw data in their reports, helping readers to pinpoint the exact studies that lead to disagreements in conclusions about an area. Third, reviewers are comparing results from unrelated meta-analyses, which will eventually lead to a better understanding of the factors influencing the outcomes of educational research. And finally, some of the worst abuses that have taken place in meta-analysis have now been left behind. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247270

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J.A. Kulik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Uses and Misuses of Meta-Analysis

James A. Kulik

The University of Michigan

An invited address at the annual meeting of the
American Educational Research Association,
New Orleans, April 1984

TM 840 443

Interest in meta-analysis has grown enormously since Glass first introduced the method in his 1976 AERA presidential address. Since that time, approximately 1000 papers on meta-analysis and quantitative research synthesis have appeared in books, journal articles, ERIC documents, and dissertations.[1] Approximately one-third of these reports describe results from actual meta-analytic studies. By almost any standard, the record shows that Glass's address has had an enormous impact on educational research.

All this should not come as a surprise to anyone who is interested in the topic of research synthesis. Most educational researchers know that meta-analysis is a hot topic, and they do not need to hear statistics to get the point. Nor is there any mystery about why meta-analysis has attracted so much attention. For a long time before Glass introduced meta-analysis, researchers felt uneasy about the reliability of qualitative research reviews in education. But no real alternative to this type of review was available.

In an article in the Review of Educational Research a few years ago, Gregg Jackson (1980) listed some of the major complaints about qualitative reviews:

1. Qualitative reviews frequently fail to cite earlier reviews on the same or similar topics.

2. They often report results from only a few studies in areas where many studies are available.

3. They frequently use crude and misleading representations of the findings of studies.

4. They sometimes fail to recognize that random sampling error contributes to variation in study outcomes.

5. They seldom examine relationships between characteristics of studies and study findings.

---

[1]These estimates are based on two sources: (a) a comprehensive bibliography on meta-analysis compiled by William K. Lamb and Dean K. Whitla of Harvard University; and (b) my recent computer search of Dialog's ERIC database for publications that use the term "meta-analysis" or "meta-analytic" in their titles or abstracts. Lamb and Whitla's comprehensive bibliography lists 261 publications that appeared before January 1981. My year-by-year search of the ERIC system suggests that approximately 75% of works related to meta-analysis were published after January 1981. It seems reasonable therefore to believe that a comprehensive bibliography through March 1984 would contain approximately 1000 references.

6. They usually report little about the methods used in the review so that readers cannot judge the validity of the conclusions.

Meta-analysis does not solve all these problems, but it holds out some hope for increasing the reliability and dependability of a reviewer's conclusions in reviews. It promises to make reviews more useful. It is hard for me to imagine therefore that there will be any turning away from meta-analysis in the future. It has the look of a survivor.

Uses of Meta-Analysis

A number of developments in the meta-analytic literature make me especially optimistic about its future. They are the kinds of things that convince me that the approach is growing better and stronger with time. This afternoon I would like to mention three of the most promising developments.

Emphasis on Replication of Meta-Analyses

First of all, it's become clear recently that meta-analyses spawn new meta-analyses in the same areas. That may not seem like much, but I find it very heartening. After Glass introduced the approach in 1976, skeptics began warning that meta-analysis might be the death of primary research. The meta-analyst would analyze the collected results in an area of research vitality, and interest in doing primary research in the area would dry up. The average Effect Size could serve as the epitaph for the area.

Just the opposite, in fact, has happened. Meta-analysis has challenged researchers to pay more attention to the areas in which it's been applied.

1. At least three independent meta-analyses, for example, have been carried out on effects of psychotherapy, and others are in progress (Landman & Dawes, 1982; Shapiro & Shapiro, 1982; Smith, Glass, & Miller, 1980).

2. At least four independent meta-analyses have been published on the effects of coaching for the Scholastic Aptitude Test (Der Simonian & Laird, 1983; J. Kulik, Bangert, & Kulik, 1984; Messick & Jungeblut, 1981; Slack & Porter, 1980).

3. At least three meta-analyses have been carried out on effects of desegregation (Crain & Mahard, 1982; Krol, 1979; Wortman & Bryant, in press).

4. And numerous meta-analyses are available in the area of computer-based instruction (Burns & Bozeman, 1981; Hartley, 1978; C. Kulik, Kulik, & Bangert-Drowns, 1984;

J. Kulik, Bangert, & Williams, 1983; J. Kulik, Kulik, & Cohen, 1980).

I see no evidence that such meta-analytic activities have dried up the stream of primary research from these areas. My colleagues and I recently decided to update an earlier meta-analysis of ours on college-level computer-based instruction. We were almost overwhelmed by the number of primary studies that have appeared since our first meta-analysis was done. It is hard for us to believe therefore that meta-analysis has had negative effects on the production of primary research. If anything, meta-analytic efforts may give primary researchers a better sense of how their work fits into a large picture.

Publication of Meta-Analytic Data

A second major development in the field of meta-analysis is also very encouraging to me. In recent years, meta-analysts have been growing more open about the data that go into their syntheses. More meta-analysts today than ever before are including in their reports lists of studies, study effects, and study features.

I am especially pleased about this development, because I know that the decision to go public with one's raw data is not an easy one for most meta-analysts to make. For one thing, it gives others equal ownership of the data the meta-analyst sweated over. It also decreases the number of publication outlets for manuscripts since many journals will not publish articles with tables of primary studies. It also decreases the likelihood that other meta-analyses in an area will be truly independent.

The advantages, however, more than make up for these drawbacks, and my colleagues and I at Michigan have decided to include our raw data in all our future publications. I hope that other meta-analysts will follow us on this course.

The great advantage of going public with one's data is to the reader. With meta-analytic data laid out before them, readers can see how meta-analysts draw their conclusions. There is also a great advantage to meta-analysts. With various meta-analytic data sets laid out before them, meta-analysts can usually pinpoint the factors that cause disagreements among reviewers. In the long run, this will be of enormous benefit to educational research and educational policy.

Going public with meta-analytic data has already cleared up some of the mystery surrounding reviewers' conclusions about coaching for the Scholastic Aptitude Test. Each of the four recent meta-analyses on SAT coaching reported very similar average effects. But the researchers

conducting the meta-analyses drew very different conclusions about the potential benefits of coaching. With the meta-analytic data from all studies available, it is easy to see that the differences in conclusions depended on the weight and interpretation that each meta-analyst gave to a few outlier cases.

Synthesis of Findings from Different Areas

There is a third important recent development in meta-analysis, and I want to applaud it too. Meta-analysts are now comparing meta-analytic findings from different areas to see whether certain study features have consistent effects on outcomes. As in so many other things, Glass and his colleagues have led the way in this endeavor too (Smith et al., 1980). They have concluded that variation in study findings is only modestly predictable from study characteristics. In nearly all instances, less than 25% of the variance in study results can be accounted for by the best combination of study features. According to Glass, only a few study features are consistently related to outcomes from study to study.

Since 1978 my colleagues and I have been carrying out meta-analyses of findings on instructional technology. We recently examined a set of ten of our meta-analyses for consistency in relationship of study features and outcomes (Bangert-Drowns, Kulik, & Kulik, 1984). Our conclusions were similar to Glass's. We found that relatively little of the variation in study findings was attributable to study features. Among the few features that consistently influenced study outcomes were the source of publication and the magnitude of the study. Effect sizes were usually higher in journal articles than they were in dissertations, and they were usually higher in small scale studies. Among the study features that showed little relationship to size of effect was the basic experimental design of the study. True experiments and quasi-experimental evaluations produced the same basic results.

Abuses of Meta-Analysis

Before discussing areas where progress still needs to be made, I want to say that meta-analysts are fortunate to be able to trace their work back to Smith, Glass, & Miller's (1980) study on effects of psychotherapy. That study stands as an intellectual feat of the first magnitude. It was bold in conception, painstaking in execution, and ingenious in design. Few studies in the history of educational research could equal its inventiveness.

It seems to me, however, that Glass and his associates in that study were testing the limits of how much could be done in analyzing an aggregation of study effects. No

matter what the original unit of measurement in a study, Glass and his colleagues were able to express the measurement in standard deviation units. And because they were able to express all measurements on the same scale, they were able to include a very large number of comparisons in a single statistical analysis. They were thus able to use the most sophisticated statistical tools in their analyses.

This bold approach was not without its problems, however. One obvious difficulty is the inflated N in the analysis--a sample size much larger than the number of independent observations. When a given study is represented two, three, four, or five times in a data set, it is difficult to determine the amount of error in statistics describing the set. It is even more difficult to estimate the true correlations among study features when single studies are multiply represented in a data set. The results from regression analyses on such data sets must usually be suspect.

But even more serious is the conceptual confusion that arises when different types of outcome measures are lumped together. Glass, Smith, and Miller's overall conclusion from their meta-analysis of psychotherapy was: "The average study showed a 0.68 standard deviation superiority of the treated group over the control group." A superiority of 0.68 standard deviations of what? Of palmar sweat? Self-satisfaction? Academic achievement? Job performance? Effects varied from one type of criterion to another.

Few of the meta-analysts who followed Glass were able to match his ingenuity and boldness, but almost anyone could outdo him in breaking the rules on inflating N's and defining criteria precisely. And in the early years, meta-analysts sometimes did just that. A few meta-analysts went to extremes to get the biggest possible N's for their studies. They coded effect sizes for every possible subgroup and subscale in each study. They acted as though entirely different outcome measures were completely equivalent. And they they gave marginal work the benefit of the doubt, throwing it into the hopper with work of far greater quality.

I think that the worst of these excesses are behind us. The problem of lumping together different outcomes in one analysis seems to me to be on its way out. Most meta-analysts now look separately at cognitive and affective outcomes of instruction, and some carry out a dozen separate analyses on a dozen different types of instructional outcomes. The problem of inflated N's seems to me, however, to be a more enduring legacy of the early meta-analytic work. Most meta-analysts continue to use as the N in their analyses a figure that is a good deal larger than the number

of independent studies they locate. I hope that some day meta-analysts will start giving more attention to the results of this practice.

Conclusion

In conclusion, I want to reiterate that I find several developments in meta-analysis a cause for optimism. First, I am encouraged that different meta-analysts are doing work in the same areas. This is leading to increased confidence in meta-analytic results. Second, I am very pleased that meta-analysts are beginning to include their raw data in their reports. That helps reader pinpoint the exact studies that lead to disagreements in conclusions about an area. Third, I am encouraged that reviewers are comparing results from unrelated meta-analyses. I think that this will eventually lead to a better understanding of the factors influencing the outcomes of educational research. And finally, I think that some of the worst abuses that have taken place in meta-analysis are behind us.

References

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1984, August). The influence of study features on outcomes of educational research. A paper to be presented at the annual meeting of the American Psychological Association, Toronto.

Burns, P. K., & Bozeman, W. C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship? Educational Technology, 21, 32-39.

Crain, R. L., & Mahard, R. E. (1982, June). Desegregation plans that raise black achievement: A review of the research. Santa Monica, CA: The Rand Corporation (N-1844-NIE).

DerSimonian, R., & Laird, N. M. (1983). Estimating the effect of coaching on SAT scores: A meta-analysis. Harvard Educational Review, 53, 1-15.

Glass, G. V, McGaw, B., & Smith, M. L. Meta-analysis in social research. Beverly Hills: Sage, 1981.

Hartley, S. S. (1978). Meta-analysis of the effects of individually paced instruction in mathematics. Dissertation Abstracts International, 38(7-A), 4003. (University Microfilms No. 77-29,926)

Jackson, G. B. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

Krol, R. A. (1979). A meta-analysis of comparative research on the effects of desegregation on academic achievement. (University Microfilms No. 7907962)

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1984). Effects of computer-based education on elementary school students. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Kulik, J. A., Bangert, R. L., & Kulik, C.-L. C. (1984). Effects of coaching for aptitude tests. Psychological Bulletin, 95, 179-188.

Kulik, J. A., Bangert, R. L., & Williams, G. W. (1983). Effects of computer-based teaching on secondary school students. Journal of Educational Psychology, 53, 143-158.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1980). Effectiveness of computer-based college teaching: A